**Professional Summary:**

* Overall 6 years of experience in design, development and implementations of robust technology systems, with specialized expertise in Hadoop Administration and Linux Administration.
* Able to understand business and technical requirements quickly; Excellent communications skills and work ethics; Able to work independently.
* 4 years of experience in Hadoop Administration & Big Data Technologies and 2 years of experience into Linux administration.
* Experience with complete Software Design Lifecycle including design, development, testing and implementation of moderate to advanced complex systems.
* Hands on experience in installation, configuration, supporting and managing Hadoop Clusters using Horton works, Cloudera.
* Hadoop Cluster capacity planning, performance tuning, cluster Monitoring, Troubleshooting.
* Design Big Data solutions for traditional enterprise businesses.
* Backup configuration and Recovery from a Name Node failure.
* Excellent command in creating Backups & Recovery and Disaster recovery procedures and Implementing BACKUP and RECOVERY strategies for off-line and on-line Backups.
* Involved in bench marking Hadoop/HBase cluster file systems various batch jobs and workloads.
* Making Hadoop cluster ready for development team working on POCs.
* Experience in minor and major upgrades of Hadoop and Hadoop eco system.
* Experience monitoring and troubleshooting issues with Linux memory, CPU, OS, storage and network.
* Hands on experience in analyzing Log files for Hadoop and eco system services and finding root cause.
* Experience on Commissioning, Decommissioning, Balancing and Managing Nodes and tuning server for optimal performance of the cluster.
* As an admin involved in Cluster maintenance, trouble shooting, Monitoring and followed proper backup& Recovery strategies.
* Good Experience in setting up the Linux environments, Password less SSH, creating file systems, disabling firewalls, swappiness, Selinux and installing Java.
* Good Experience in Planning, Installing and Configuring Hadoop Cluster in Cloudera and Hortonworks Distributions.
* Installing and configuring Hadoop eco system like pig, hive.
* Hands on experience in Installing, Configuring and managing the Hue and HCatalog.
* Experience in importing and exporting the data using Sqoop from HDFS to Relational Database systems/mainframe and vice-versa.
* Experience in importing and exporting the logs using Flume.
* Optimizing performance of Hbase/Hive/Pig jobs.
* Hands on experience in Zookeeper and ZKFC in managing and configuring in NameNode failure scenarios.
* Handsome experience in Linux admin activities on RHEL & Cent OS.
* Experience in deploying Hadoop 2.0(YARN).
* Familiar with writing Oozie workflows and Job Controllers for job automation.
* Hands on experience in provisioning and managing multi-tenant Hadoop clusters on public cloud environment – Amazon Web Services (AWS)-EC2 and on private cloud infrastructure – Open Stack cloud platform.

**Education Qualifications:**

**Education:** Bachelor in Computer Science, JNTUH University, India.

**Technical skills**

|  |  |
| --- | --- |
| Hadoop | Hadoop, Pig, Hive, Tez, Sqoop, Impala, Spark, Cassandra, Sqoop, Oozie, HBase, Kafka  Cluster **Mgmt tools:** Apache Ambari, Puppet, Ansible, Ganglia, Nangios, Cloudera Manager, Horton works HDP Big Data Distributions: Apache, Cloudera(3.x,4.x,5.x), Horton works Data Platform(HDP) (2.x),Kerberos, Knox, Ranger, Sentry |
| ETL/BI Tools | Abinitio, Abinitio MDH, Tableau, SAS, Informatica Power Center 8.6, Datameer, Micro Strategy, Attunity |
| RDBMS | SQL Server, DB2, Teradata, MySQL, Oracle |
| Database | MySQL, NoSQL, Hbase, Oracle, Green plum, PostgreSQL, DB2. |
| Operating System | UNIX, Linux, Windows XP, Windows Vista, Windows 2003 Server. |
| Learning | Cloud clustering, AWS, Rack space |
| Scripting Languages: | JavaScript, Python, Shell. |

**Professional Experience:**

**Client:** Ameren St. Louis, MO. From: Sep’18 to Till Date.

**Role:** Sr Hadoop Administrator.

**Responsibilities:**

* Deployed multi-node development, testing and production Hadoop clusters with different Hadoop components **(**HIVE, PIG, SQOOP, OOZIE, FLUME, HCATALOG, ZOOKEEPER) using Hortonworks Ambari.
* Configured Capacity Scheduler on the Resource Manager to provide a way to share large cluster resources.
* Deployed Name Node high availability for major production cluster and designed automatic failover control using zookeeper and quorum journal nodes.
* Configured Oozie for workflow automation and coordination.
* Good experience in troubleshoot production level issues in the cluster and its functionality.
* Backed up data on regular basis to a remote cluster using distcp.
* Implemented High Availability and automatic failover infrastructure to overcome single point of failure for Name node utilizing Zookeeper services.
* Regular Ad-Hoc execution of Hive and Pig queries depending upon the use cases.
* Regular Commissioning and Decommissioning of nodes depending upon the amount of data.
* Experience in Disaster Recovery and High Availability of Hadoop clusters/components.
* Monitor Hadoop cluster connectivity and security. Manage and review Hadoop log files.
* File system management and monitoring. HDFS support and maintenance.
* Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability.
* Monitored Hadoop Jobs and Reviewed Logs of the failed jobs to debug the issues based on the errors.
* Optimized Hadoop clusters components: HDFS, Yarn, Hive, Kafka to achieve high performance.
* Worked with Linux server admin team in administering the server Hardware and operating system.
* Interacted with Networking team to improve bandwidth.
* Integrated Oozie with the rest of the Hadoop stack supporting several types of Hadoop jobs such as MapReduce, Pig, Hive, and Sqoop as well as system specific jobs such as Java programs and Shell scripts.
* Installed Kafka cluster with separate nodes for brokers.
* Diagnose and resolve performance issues and scheduling of jobs using
* Configured Fair scheduler to share the resources of the cluster.
* Experience designing data queries against data in the HDFS environment using tools such as Apache Hive.
* Imported data from MySQL server to HDFS using Sqoop.
* Manage the day-to-day operations of the cluster for backup and support.
* Used the RegEx, JSON and Avro SerDe’s for serialization and de-serialization packaged with Hive to parse the contents of streamed log data.
* Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce, Hive and Sqoop as well as system specific jobs.
* Worked along with the Hadoop Operations team in Hadoop cluster planning, installation, maintenance, monitoring and upgrades.

**Environment**: Ambari, Hbase, Hive, Pig, Sqoop, Apache Ranger, Splunk, Yarn, Apache Oozie workflow scheduler, Flume, Zookeeper, RegEx, JSON.

**Client:** General Motors, Austin, TX. **From:** July’17 – Aug ‘18

**Role:** Sr. Hadoop Administrator.

**Responsibilities:**

* Setup a Multi Node Cluster. Plan and Deploy a Hadoop Cluster using Hortonworks Ambari.
* Secure a deployment and understand Backup and Recovery.
* Performed in developing purge/archive criteria and procedures for historical.
* Performance tuning of Hadoop clusters and Hadoop MapReduce routines.
* Screen Hadoop cluster job performances and capacity planning
* Monitor Hadoop cluster connectivity and security. Manage and review Hadoop log files.
* File system management and monitoring. HDFS support and maintenance.
* Diligently teaming with the infrastructure, network, database, application and business intelligence teams to guarantee high data quality and availability.
* Collaborating with application teams to install operating system and Hadoop updates, patches, version upgrades when required. Point of Contact for Vendor escalation.
* Created reports for the BI team using Sqoop to export data into HDFS and Hive.
* Used Sqoop for bringing in the raw data, populate staging tables and store the refined data in partitioned tables.
* Created Hive queries that helped market analysts spot emerging trends by comparing fresh data.
* Developed Hive queries to process the data for analysis by imposing read only structure on the stream data.
* Performed minor and major upgrades, commissioning and decommissioning of data nodes on Hadoop cluster.
* Installed Hadoop eco system components like Pig, Hive, Hbase and Sqoop in a Cluster.
* Experience in setting up tools like Ganglia for monitoring Hadoop cluster.
* Handling the data movement between HDFS and different web sources using Flume and Sqoop.
* Extracted files from NoSQL database like HBase through Sqoop and placed in HDFS for processing.
* Installed Oozie workflow engine to run multiple Hive and Pig jobs.
* Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase.
* The Hive tables created as per requirement were internal or external tables defined with proper static and dynamic partitions, intended for efficiency.
* Used Oozie workflow engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce, Hive and Sqoop as well as system specific jobs.
* Worked along with the Hadoop Operations team in Hadoop cluster planning, installation, maintenance, monitoring and upgrades.
* Adding, removing, or updating user account information, resetting passwords, etc.
* Installing and updating packages using YUM.

**Environment:** Ambari, Multi-Node cluster Set up, HIVE, PIG, SQOOP, Mrv2, YARN Framework and Apache Oozie workflow scheduler, Apache Ranger, Splunk.

**Client:** Mattel Inc, El Segundo, CA. **From:** Nov’16 – June’17

**Role:** Hadoop Administrator

**Responsibilities:**

* Involved in monitoring many Hadoop clusters using CDH5.4.
* Working on CDH4, CDH5 including High Availability, YARN, data streaming, security, application deployment
* Worked on capacity planning for growing Hadoop clusters.
* Used to develop data pipelines that ingest data from multiple data sources and process them.
* Expertise in Using Sqoop to connect to the ORACLE, DB2 and move the pivoted data to Hive tables or Avro files.
* Developed Simple to complex MapReduce Jobs using Hive and Pig
* Involved in creating Hive tables, and loading and analyzing data using hive queries
* Extensively used Sqoop to move the data from relational databases to HDFS.
* Used Flume to move the data from web logs onto HDFS.
* Used Pig to apply transformations, cleaning and duplication of data from raw data sources.
* Upgraded the Hadoop cluster from CDH4 to CDH5.
* Configured and deployed hive Meta store using MySQL and thrift server.
* Launching and Setup of HADOOP Cluster on AWS, which includes configuring different components of HADOOP.
* Deployed high availability on the Hadoop cluster using quorum journal nodes.
* Implemented automatic failover zookeeper and zookeeper failover controller.
* Commissioning and Decommissioning of nodes depending upon the amount of data.
* Integrated all the clusters Kerberos with Company’s Active Directory, and created USERGROUPS and PERMISSIONS for authorized access in to the cluster.
* Configured Ganglia which include installing gmond and gmetad daemons which collects all the metrics running on the distributed cluster and presents them in real-time dynamic web pages which would further help in debugging and maintenance.
* Involved in deploying a Hadoop cluster using CDH4 integrated with Nagios.
* Responsible for building system that ingests terabytes of data per day into Hadoop from a variety of data sources providing high storage efficiency and optimized layout for analytics.
* Used to develop data pipelines that ingest data from multiple data sources and process them.
* Performed Installation and configuration of Hadoop Cluster of 90 Nodes with Cloudera distribution with CDH4.
* Created a local YUM repository for installing and updating packages.
* Performed various configurations, which includes, networking and IPTable, resolving hostnames, user accounts and file permissions, http, ftp, SSH keyless login.

**Environment**: Hadoop Hdfs, Mapreduce, Hive, Pig, Flume, Oozie, Sqoop, Eclipse, Hortonworks, Ambari.

**Client:** BCG Platinion, Pittsburg, PA. **From:** Jan’15 - Oct’16

**Role:** Hadoop Administrator

**Responsibilities:**

* Hadoop installation, Configuration of multiple nodes using Cloudera platform.
* Major and Minor upgrades and patch updates.
* Handling the installation and configuration of a Hadoop cluster.
* Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase.
* Handling the data exchange between HDFS and different web sources using Flume and Sqoop.
* Monitoring the data streaming between web sources and HDFS.
* Monitoring the Hadoop cluster functioning through monitoring tools.
* Close monitoring and analysis of the MapReduce job executions on cluster at task level.
* Inputs to development regarding the efficient utilization of resources like memory and CPU utilization based on the running statistics of Map and Reduce tasks.
* Changes to the configuration properties of the cluster based on volume of the data being processed and performed by the cluster.
* Setting up automated processes to analyze the system and Hadoop log files for predefined errors and send alerts to appropriate groups.
* Excellent working knowledge on SQL with databases.
* Commissioning and De-commissioning of data nodes from cluster in case of problems.
* Setting up automated processes to archive/clean the unwanted data on the cluster, in particular on Name Node and Secondary Name node.
* Set up and managing HA Name Node to avoid single point of failures in large clusters.
* Discussions with other technical teams on regular basis regarding upgrades, process changes, any special processing and feedback.

**Environment**: Java, Linux, Shell Scripting, Teradata, SQL server, Cloudera Hadoop, Flume, Sqoop, Pig, Hive, Zookeeper and HBase.

**Client:** HTC Global Services India. **From:** Mar’14 – Nov’14

**Role**: LINUX Administrator.

**Responsibilities:**

* Installed, configured, maintained and administrated the network servers DNS, NFS and application servers Apache and Samba server.
* Installation and Configuration of SSH, TELNET, FTP, DHCP, DNS.
* Worked on UNIX shell scripting for system/application in automating server task, installing and monitoring applications and data feeding file transfer and log files.
* Maintained UNIX (Red Hat Enterprise Linux4, 5, CentOS4, 5, VMware) on Sun
* Enterprise servers & Dell Servers.
* Implemented the Jumpstart servers and Kick start Servers to automate the server builds for multiple profiles.
* Installed and deployed RPM Packages.
* APACHE Server Administration with Virtual Hosting.
* Worked extensively in using VI editor to edit necessary files writing shell script.
* Worked on adding new Users and +groups and give sudo access and central file synchronization via sudoers, authorized keys, password, shadow, and group.
* Coordinated with application team in installation, configuration and troubleshoot issues with Apache, Web logic on Linux servers.
* Local and Remote administering of servers, routers and networks using Telnet and SSH.
* Monitored client disk quotas &disk space usage.
* Worked on backup technologies like VeritasNetbackup4.x, 5.0, 6.x and Tivoli Storage Manager 5.5.
* Involved in back up, firewall rules, LVM configuration, monitoring servers and on call support.
* Created BASH shell scripts to automate cron jobs and system maintenance. Scheduled cron jobs for job automation.

**Environment**: Red hat 4/5, Solaris 8/9/10, CentOS 4/5, SUSE Linux 10.1/10.3, VMware

**Client:** Genesis Tech solutions, India **. From:** May’13–Feb’14

**Role:** LINUX Administrator.

**Responsibilities**:

* Installation, Configuration, Up gradation and administration of Windows, Sun Solaris, Redhat Linux and Solaris.
* Linux and Solaris installation, administration and maintenance.
* User account management, managing passwords setting up quotas and support.
* Worked on Linux Kick-start OS integration, DDNS, DHCP, SMTP, Samba, NFS, FTP, SSH, and LDAP integration.
* Network traffic control, IPsec, Quos, VLAN, Proxy, Radius integration on Cisco Hardware via Red Hat Linux Software.
* Installation and configuration of MySQL on Windows Server nodes.
* Responsible for configuring and managing Squid server in Linux and Windows.
* Configuration and Administration of NIS environment.
* Package and Patch management on Linux servers.
* Worked on Logical volume manager to create file systems as per user and database requirements.
* Data migration at Host level using Red Hat LVM, Solaris LVM, and Veritas Volume Manager.
* Expertise in establishing and documenting the procedures to ensure data integrity including system fail-over and backup/recovery in AIX operating system.
* Managed 100 + UNIX servers running RHEL, HPUX on Oracle HP.
* Solaris Disk Mirroring (SVM), ZONE installation and configuration
* Escalating issues accordingly, managing team efficiently to achieve desired goals.

**Environment**: Linux, TCP/IP, LVM, RAID, Networking, Security, user management, MySQL.